VACCINE

FIELD OF INVENTION

This invention relates to a general method for detecting pathogenic strains of bacteria that harbour a type III secretion system, and characterising regions of the chromosome of said strain where virulence genes reside. More particularly, this invention relates to the method as applied to the pathogen *Bordetella pertussis*. Furthermore, the invention relates to newly identified polynucleotides within these regions, virulent polypeptides encoded by them and to the use of such polynucleotides and polypeptides, and to their production.

BACKGROUND OF THE INVENTION

Type III secretion systems:

Pathogenic bacteria invade many different niches in a broad host range and cause a wide variety of syndromes. It is due to this fact that it was believed previously that each disease might be induced by a distinct molecular mechanism. However, the spectrum of such mechanisms is not as broad as first imagined; rather, bacteria exploit a number of common molecular tools to achieve a range of goals. Among these tools are type III secretion systems, which provide a means for bacteria to target virulence factors directly at host cells. These factors then tamper with host cell functions to the pathogens' benefit.

The type III export system is responsible for secretion of *Salmonella* and *Shigella* invasion and virulence factors, Enteropathogenic *Escherischia coli* (EPEC) signal transduction molecules, virulence factors in several plant pathogens (for instance *Xanthomonas campestris pv. vesicatoria* [Fenselau *et al.*, 1992]) and Yops proteins in *Yersinia*. Yops export mechanism has been the most intensively investigated type III secretion apparatus (see for instance: Allaoui *et al.*, 1994; Bergman *et al.*, 1994). In this system, more than 20 different Ysc/Lcr proteins, all encoded by the virulence plasmid pYV, are presumed to compose a secretion channel spanning the *Yersinia* cell envelope.

Besides these elements involved in the secretion machinary, the pYV plasmid codes for the Yops proteins which are the secreted substrates and appear as the actual effectors of virulence.

Comparative studies of type III secretion systems originating from different species reveal that the components of the secretion machinery are conserved (Gygi *et al.*, 1995; Bogdanove *et al.*, 1996). In addition, homologs have been found in determinants which take part in flagellar assembly, indicating that this secretion pathway may be involved in surface organelle biosynthesis (Ramakrishnan *et al.*, 1991).

In contrast, however, the secreted substrates share no similarities, except in few cases. Therefore, the abandoned concept of a distinct molecular mechanism corresponding to each disease could reappear at the level of effector proteins.

Pathogenicity island

Pathogenicity islands have emerged as a novel theme in the field of bacterial virulence. Although they can comprise type III secretion systems they do not exlusively do so.

Early in the search for virulence genes, it was observed that many of these genes resided on plasmids. However, numerous virulence genes were also found on the chromosome. Surprisingly, the chromosomal virulence genes are also often clustered in functionally related groups. Such groups of virulence genes gave rise to the concept of pathogenicity islands (Pais) which can be defined as compact, distinct genetic units carrying virulence genes. These units, often flanked by direct repeats, occupy large chromosomal regions (often > 30 kb) and are present in pathogenic strains, whilst being absent or sporadically distributed in less-pathogenic (or non-pathogenic) strains of a bacterial species. These DNA segments are frequently associated with tRNA genes

and/or insertion sequence (IS) elements at their boundaries. In addition, their G+C content often differs from that of host bacterial DNA, suggesting a foreign origin.

Pathogenicity islands have been discovered in an increasing number of bacterial pathogens, including different categories of *E. coli, Salmonella typhimurium, Yersinia* spp, *Helicobacter pylori*, *Vibrio cholera* etc.

The first intensively studied pathogenicity islands were Pai I and Pai II, which encode the haemolysin determinants of uropathogenic *E. coli*. These two Pais, are flanked by direct repeats and can be deleted from the chromosome at frequencies of 10^{-4}, resulting in non-virulent mutant strains. Another pathogenicity island of 35 kb has recently been identified on the chromosome of enteropathogenic *E. coli* (EPEC) and was found to encode all known determinants involved in the so-called "attaching and effacing" (AE) lesion formation. This region was therefore referred to as "locus of enterocyte effacing" (LEE). Despite the fact that uropathogenic and enteropathogenic *E. coli* cause completely different infectious diseases, Pai I of the uropathogenic strains and the LEE locus of EPEC are inserted at exactly the same positions into the *E. coli* chromosome.

While some authors support a definition of pathogenicity islands which necessarily includes its chromosomal location, others have extended the concept to blocks of virulence genes, regardless of their location in chromosomes, plasmids or phages. The fact that, on one hand, phages and plasmids can easily insert into and excise from the chromosome and, on the other, that cryptic origins of plasmid replication, or phage related sequences were detected in Pais, prompted the latter and less restrictive definition.

The pathogenicity islands (PAIs) which code for a type III secretion system encompass genes that divide into two classes, I and II. Class I encompasses the genes coding for the secretion machinery components and their regulators of expression, class

II encompasses the genes encoding secreted effector proteins. Both *Yersinia lcrD* and *yscU* belong to class I. The precise functions of class I determinants is not well understood. Although it is sometimes not straightforward to make a clear distinction between class I and class II components, genes of class I can be identified as being present in many different species, and a comparison of their respective gene sequences indicate that equivalent genes share a significant (*yscI, yscO*) or even high level (*lcrD, yscU, yscN*) of sequence similarity (Hueck, 1998).

The second class of genes (class II) codes for proteins which constitute the substrate secreted by the translocon. These proteins appear as the actual effectors of virulence and are referred to as target proteins, virulence effector proteins or, simply, effectors. In contrast to the situation prevailing in class I gene products, the effectors share no, or very weak, similarities between species. Effector proteins are those which present the best biological, vaccine and diagnostic potentialities.

The inventors have discovered that the clustering of class I and class II genes inside a single pathogenicity island, offers the opportunity of conveniently finding and characterising unknown class II genes by targeting class I genes which can be identified using a known sequence of one of their numerous orthologues.

Bordetella pertussis

Whooping cough is a disease caused by infection by *Bordetella pertussis*, and is a serious and debilitating human disease particularly in young children. Although whole cell and acellular vaccines are available that are effective against the disease, there remains a need for the identification of further highly purified pertussis proteins that could be used in a more efficacious pertussis vaccine.

Although many pertussis virulence associated factors are known such as pertussis toxin, filamentous haemagglutinin, pertactin, which have been included in various acellular vaccines, there is no convenient genetic method for identifying further virulence factors using the pertussis genome (short of laboriously sequencing the whole genome). Although class I type III secretion system virulence genes have recently been shown to exist in *B. bronchiseptica* and *B. pertussis* (Yuk *et al.*, 1998), there has been no complete analysis of a pathogenicity island in *Bordetella*, and the identity and characterisation of effector genes within such a pathogenicity island has been unknown up until the present invention.

SUMMARY OF THE INVENTION

In one aspect, the invention relates to a method for the identification of new virulence genes in bacterial strains containing a type III secretion system. In particular, the invention allows the identification of the effector virulence genes associated within a pathogenicity island containing the genes for the type III secretion system. Another aspect of the invention a method for the identification of pathogenic bacterial strains containing a type III secretion system. Another aspect of the invention relates to *Bordetella pertussis* BopN, Orf1, Orf2, Orf3, Orf4, Orf5, Orf6, Orf7, Orf8, Orf9, Orf10, Orf11, Orf12, Orf13, Orf14, Orf15 effector proteins, and the respective polynucleotide sequences encoding them.

Although the general concepts of type III secretion systems and pathogenicity islands have been reported, the problem of how simply and reliably to identify whether any given organism has such cell machinery has not been accomplished until now. Such a method is extremely useful to establish whether a given strain has a type III secretion system within a pathogenicity island, to characterise unknown virulence genes within the pathogenicity island, and to use in quick diagnostic methods for determining whether a cultured bacterial strain containing a type III secretion system is pathogenic.

In the present invention, a novel, general method is described to achieve the above aims. More specifically, the invention utilises a method that employs ideally-suited primers designed specifically from the sequence of the virulent *Yersinia enterocolitica lcrD* gene as a target sequence. The presence of a type III secretion system within a pathogenicity island in *Bordetella pertussis* was discovered, and every gene within the pathogenicity island was characterised.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Nucleotide and deduced amino acid sequences of the cloned 152 bp amplicon. The primers involved in the original amplification, the subsequent nested PCR, and the gene library screening are all derived from this sequence, and listed specifically in Table 1.

Fig. 2. PileUp figure from the deduced amino acid sequences homologous to *Yersinia* LcrD. Abbreviations: BbuFlhA = *Borrelia burgdorferi* FlhA; TpaFlhA = *Treponema pallidum* FlhA; BsuFlhA = *Bacillus subtilis* FlhA; CjeFlbA = *Campylobacter jejuni* FlbA; HpyFlhA = *Helicobacter pylori* FlhA; EcoFlhA - *Escherichia coli* FlhA; StyFlhA = *Salmonella typhimurium* FlhA; YenFlhA = *Yersinia enterocolitica* FlhA; PmiFlhA = *Proteus mirabilis* FlhA; CcrFlbF = *Caulobacter crescentus* FlbF; EcoFhiA = *Escherichia coli* FhiA; EamHrpI = *Erwinia amylovora* HrpI; PsyHrpI = *Pseudomonas syringae* HrpI; ECEPSepA = Enteropathogenic *Escherichia coli* SepA; StySsaV = *Salmonella typhimurium* SsaV; RsoHrpO = *Ralstonia solanacearum* HrpO; XcaHrpC2 = *Xanthomonas campestris* HrpC2; SflMxiA = *Shigella Flexneri* MxiA; StyInvA = *Salmonella typhimurium* InvA; PaePcrD = *Pseudomonas aeruginosa* PcrD; YenLcrD = *Yersinia enterocolitica* LcrD; BpeBcrD = *Bordetella pertussis* BcrD; CpsTtsB = *Chlamydia psittaci* TtsB.

Fig. 3. Organization of the *Bordetella pertussis* pathogenicity island (Pai). Four house keeping genes (hatched boxes) and the transposase gene of IS*481* (black box) are surrounding the Pai. The Pai consists of genes coding for determinants involved in the secretory apparatus and its regulation (class I genes, in grey boxes) as well as ORFs which putitively code for effector proteins (class II genes, in white boxes). Letters indicate the respective class I *bsc* genes whereas numbers correspond to the class II ORFs listed in Table 3.

Fig. 4. PileUp figure from the deduced amino acid sequences homologous to *Yersinia* YscU. Abbreviations: BbuFlhB = *Borrelia burgdorferi* FlhB; TpaFlhB = *Treponema pallidum* FlhB; EcoFlhB - *Escherichia coli* FlhB; StyFlhB = *Salmonella typhimurium* FlhB; PmiFlhBpart = partial *Proteus mirabilis* FlhB; YenFlhB = *Yersinia enterocolitica* FlhB; BsuFlhB = *Bacillus subtilis* FlhB; HpyFlhB = *Helicobacter pylori* FlhB; AtuFlhB = *Agrobacterium tumefaciens* FlhB; CcrPodW = *Caulobacter crescentus* PodW; SflSpa40 = *Shigella flexneri* Spa40; StySpaS = *Salmonella typhimurium* SpaS; EcoEscU = *Escherichia coli* EscU; StySsaU = *Salmonella typhimurium* SsaU; BpeBscU = *Bordetella pertussis* BscU; YenYscU = *Yersinia enterocolitica* YscU; RsoHrpN = *Ralstonia solanacearum* HrpN; XcaOrf0part = partial *Xanthomonas campestris* Orf0; EamHrcU = *Erwinia amylovora* HrcU; EheHrcUpart = partial *Erwinia herbicola* HrcU; PsyHrpY = *Pseudomonas syringae* HrpY; CpsOrf1 = *Chlamydia psittaci* Orf1.

Fig. 5. The DNA sequence of the Bordatella pertussis genome comprising the type III secretion system pathogenicity island. Reference should be made to tables 2, 3, and 4 and Fig. 3 for information regarding open reading frames.

Fig. 6. Purification of MBP-Orf2, -4, -6 and –10 by affinity chromatography. The ultracentrifugatrion supernatants of each lysate (left part of the panels) and the products eluated from the affinity column (right part of the panels) were analysed by SDS-PAGE and revealed by Coomassie blue staining.

DESCRIPTION OF THE INVENTION

Type III secretion systems identified to date are encoded by either chromosal or plasmidic pathogenicity island genes. However, no where in the prior art was it realised that the conservation of genes encoding class I components of type III secretion systems and the clustering of these genes with effector protein coding sequences offered the opportunity for detecting unidentified target proteins involved in host colonisation. Such proteins would be potentially valuable in both vaccinal and diagnostic fields.

Although the known sequence of a gene encoding any conserved (class I) type III secretion machinery protein can be used in performing this invention, the *lcrD* gene is preferred. The chosen gene will act as a target for detecting unidentified pathogenicity islands in related bacterial species. The *lcrD* gene from *Yersinia* is preferred as it codes for the archetype of the recently identified LcrD/FlbF family of proteins. Members of this family are involved in host cell invasion, virulence in several phytopathogenic bacteria or in flagellar assembly. *lcrD* is preferred because the LcrD protein, and consequently the gene encoding it, is one of the most conserved determinants of the secretion machinary. Additionally, multiple amino acid comparisons have shown that the classification of the LcrD family members can be split into two main subfamilies, which, interestingly, can be correlated with the functions assigned to these proteins of each subfamily. One subfamily encompasses all the motility-involved proteins, while the other encompasses all the virulence-related determinants. This observation is illustrated in Fig. 2 (and mentioned in Gyri *et al.* (1995) & Bogdanove *et al.* (1996)). Thus, if an unknown *lcrD* homologous gene is identified, it may, after being routinely sequenced, be classified as a virulence or a flagellar gene. Once the pathogenicity island is identified, this simple test would therefore define whether the search for other virulence genes on the pathogenicity island should be initiated.

The preferred method for identifying unknown pathogenicity islands comprising a type III secretion system is by:

i) identifying two highly conserved regions of the target protein sequence (preferably of LcrD). Preferably, both regions should contain conserved amino acids which are encoded by the fewest number of codon possibilities e.g. Methionine (ATG being the only possibility) or Tryptophan (TGG being the only possibility). This minimises the number of permutations in both degenerate primer sets that are designed in the next stage of the process, thus ensuring a greater probability that each primer set will specifically anneal to the unknown *lcrD*-equivalent gene (thereby minimising background non-specific interactions). Most preferably, regions should also be chosen that are clearly distinguishable from the paralogue *flhA* flagellar genes, present in all flagellated bacterial strains.

ii) designing a degenerate set of primers for both of the chosen regions such that a) the primers are at least 15 bases long, preferably 20-30 bases long, and still more preferably 21-23 bases long, b) they are degenerate at bases that can be more than one type of nucleotide whilst still encoding the same amino acid (due to the degeneracy of codon usage for amino acids), but no more degenerate than is required to cover all permutations for the amino acid region selected, and c) the primer set that encodes the more N-terminal region of the chosen protein should correspond to the coding strand of its corresponding double-stranded DNA sequence, and the set that encodes the more C-terminal region should correspond to the complementary strand of the corresponding double-stranded DNA sequence.

iii) synthesising the degenerate primer sets of step ii) using conventional DNA synthesis methods well known in the art.

iv) purifying the primer sets of step iii)

v) adding both the primer sets and a sample containing nucleic acid from a bacterial strain (preferably a cell sample of the bacterial species itself) together in appropriate quantities and in an appropriate buffer in order to perform a polymerase chain reaction (PCR)

vi) performing a PCR reaction in order to amplify the region of the gene between the two primers (conditions for performing the PCR reaction can be optimised using techniques well known in the art)

vii) observing the reaction products on a gel (preferably an agarose gel) for an amplified product of the size expected; if no such product is present, the bacterial strain is unlikely to use a type III secretion system; if such a product is present, the bacterial strain is likely to have a type III secretion system, and is likely to be pathogenic.

The preferred method for confirming that the amplified product actually corresponds to a virulence gene is by carrying out steps i)-vii) above (where the target protein is LcrD) and then:

viii) optionally separating the product of correct size from any background products of incorrect size by removing the correct band from the gel, purifying the product by conventional means, and amplifying the product once more with the two degenerate primer sets in another PCR reaction (under preferably more stringent PCR conditions) [this step is required should the product of step vii) not be pure enough for direct cloning]

ix) inserting the DNA fragment by conventional means into a vector which is capable of being sequenced, and sequencing the fragment

x) comparing the deduced amino acid sequence of ix) with that of known members of the LcrD/FlbF family of proteins to associate the amplified product as being part of either a virulence or a flagellar gene.

And optionally:

xi) using the internal sequence of the fragment to design primers that are the exact sequence of, and specific to, the unknown *lcrD*-equivalent gene.

xii) using the primers of xi) firstly to screen a genomic library of the organism for positive clones

xiii) isolating the clones of xii), and sequence one or more of said clones

xiv) scanning the sequence of one clone (and overlapping sequences of other clones) to search for an open reading frame which is approximately the same size as *lcrD* (approximately 2100bp), and encodes a protein homologous to LcrD

xv) ascertaining whether the LcrD-equivalent protein is more homologous with the *flbF* (flagellar protein secretion) gene family or the *lcrD* (type III secretion system pathogenicity island) gene family.

The preferred method for characterising the whole pathogenicity island and defining unidentified virulence effector genes is by carrying out steps i)-xv) above (where the target protein is LcrD) and then:

xvi) if the sequence is more homologous with the *lcrD* gene family, designing primers at either extreme of the gene sequence already ascertained, and scanning and sequencing the genomic library (using a standard chromosome walking strategy – where the insert boundaries of an original clone serves as a probe for screening and cloning adjacent regions) to sequence eventually the whole of the pathogenicity island (both boundaries of which will be defined by the presence of either direct or inverted repeats, or insertion sequences, or the presence of house-keeping genes)

xvii)defining unidentified virulence effector genes within the sequenced pathogenicity island

xviii)cloning, expressing and characterising the virulence genes of xvii) which encode virulence effector proteins of the organism

Definitions

"Bordetella pathogenicity proteins" refers generally to polypeptides having the amino acid sequence encoded by the genes defined in tables 2 and 3, or an allelic variant thereof. These proteins are: BcrD, BcrH, BscC, BscD, BscE, BscF, BscI, BscJ, BscK, BscL, BscN, BscO, BscP, BscQ, BscR, BscS, BscT, BscU, BscV, BrpL, BopN, Orf1, Orf2, Orf3, Orf4, Orf5, Orf6, Orf7, Orf8, Orf9, Orf10, Orf11, Orf12, Orf13, Orf14, Orf15.

"Bordetella pathogenicity genes" refers to polynucleotides having the nucleotide sequence defined in tables 2 and 3, or allelic variants thereof and/or their complements. These genes are: *bcrD, bcrH, bscC, bscD, bscE, bscF, bscI, bscJ, bscK, bscL, bscN, bscO, bscP, bscQ, bscR, bscS, bscT, bscU, bscV, brpL, bopN, orf1, orf2, orf3, orf4, orf5, orf6, orf7, orf8, orf9, orf10, orf11, orf12, orf13, orf14, orf15.*

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from posttranslational natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation,

proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications has been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Variant" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential

properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions (preferably conservative), additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Variants should retain one or more of the biological activities of the reference polypeptide. For instance, they should have similar (preferably the same) antigenic or immunogenic activities as the reference polypeptide. Antigenicity can be tested using standard immunoblot experiments, preferably using polyclonal sera against the reference polypeptide. The immunogenicity can best be tested by measuring antibody responses (using polyclonal sera generated against the variant polypeptide) against purified reference polypeptide in a standard ELISA test. Preferably, a variant would retain all of the above biological activities.

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New

York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heijne, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research* (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J Molec Biol* (1990) 215:403). Most preferably, the program used to determine identity levels was the GCG 9 package, as was used in the Examples below.

As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include on average up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal

positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Polypeptides of the invention

In one aspect, the present invention relates to Bordetella pathogenicity proteins (or polypeptides). The Bordetella pathogenicity polypeptides include the polypeptides encoded by the genes defined in tables 2 and 3; as well as polypeptides comprising the amino acid sequence encoded by the genes defined in tables 2 and 3; and polypeptides comprising the amino acid sequence which have at least 75% identity to that encoded by the genes defined in tables 2 and 3 over their entire length, and preferably at least 80% identity, and more preferably at least 90% identity. Those with 95-99% identity are highly preferred.

The Bordetella pathogenicity polypeptides (or fragments thereof) may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It may be advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues or Maltose Binding Protein (MBP), or an additional sequence for stability during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

Fragments of the Bordetella pathogenicity polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the aforementioned Bordetella pathogenicity polypeptides. As with Bordetella pathogenicity polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number

1-20, 21-40, 41-60, 61-80, 81-100, and 101 to the end of Bordetella pathogenicity polypeptide. In this context "about" includes the particularly recited ranges larger or smaller by several, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes. The fragments should comprise at least 7 consecutive amino acids from the sequences e.g. 8, 10, 12, 14, 18, 20 or more depending on the particular sequence). Preferably the fragments comprise an epitope from the sequence.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of Bordetella pathogenicity polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus and/or transmembrane region or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Other preferred fragments are biologically active fragments. Biologically active fragments are those that mediate Bordetella pathogenicity protein activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Preferably, all of these polypeptide fragments retain the biological activity (for instance antigenic or immunogenic) of the Bordetella pathogenicity protein, including antigenic activity. Variants of the defined sequence and fragments also form part of the present invention. Preferred variants are those that vary from the referents by conservative amino acid substitutions i.e., those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic

residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination. Most preferred variants are naturally occurring allelic variants of Bordetella pathogenicity polypeptide present in strains of *Bordetella pertussis*.

The proteins may be chemically conjugated, or expressed as recombinant fusion proteins allowing increased levels to be produced in an expression system as compared to non-fused protein. The fusion partner may assist in providing T helper epitopes (immunological fusion partner), preferably T helper epitopes recognised by humans, or assist in expressing the protein (expression enhancer) at higher yields than the native recombinant protein. Preferably the fusion partner will be both an immunological fusion partner and expression enhancing partner.

The Bordetella pathogenicity polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

It is most preferred that a polypeptide of the invention is derived from *Bordetella pertussis*, however, it may preferably be obtained from other organisms of the same taxonomic genus. A polypeptide of the invention may also be obtained, for example, from organisms of the same taxonomic family or order, such as *Bordetella parapertussis* or *Bordetella bronchiseptica*.

A further aspect of the invention is substantially purified Bordetella pathogenicity polypeptides of the invention. "substantially purified" when used in reference to a protein or peptide means that the molecule has been largely, but not necessarily wholly, separated an purified from other cellular and non-cellular components. Typically a protein is substantially pure when it is at least about 60 % by weight free from other

naturally occurring organic molecules. Preferably the purity is at least about 75 %, more preferably at least about 90% , and most preferably at least about 99% by weight pure.

Polynucleotides of the invention

Another aspect of the invention relates to Bordetella pathogenicity polynucleotides. Bordetella pathogenicity polynucleotides include isolated polynucleotides which encode the Bordetella pathogenicity polypeptides and fragments respectively, and polynucleotides closely related thereto or variants thereof. More specifically, Bordetella pathogenicity polynucleotides of the invention include a polynucleotide comprising the nucleotide sequence of genes defined in table 2 or 3, encoding a Bordetella pathogenicity polypeptide. Bordetella pathogenicity polynucleotides further include a polynucleotide comprising a nucleotide sequence that has at least 75% identity over its entire length to a nucleotide sequence encoding the Bordetella pathogenicity polypeptide encoded by the genes defined in tables 2 and 3, and a polynucleotide comprising a nucleotide sequence that is at least 75% identical to that of the genes defined in tables 2 and 3. In this regard, polynucleotides at least 80% identical are particularly preferred, and those with at least 90% are especially preferred. Furthermore, those with at least 95% are highly preferred and those with at least 98-99% are most highly preferred, with at least 99% being the most preferred. Also included under Bordetella pathogenicity polynucleotides is a nucleotide sequence which has sufficient identity to a nucleotide sequence of a gene defined in tables 2 and 3 to hybridize under conditions useable for amplification or for use as a probe or marker. The invention also provides polynucleotides which are complementary to such Bordetella pathogenicity polynucleotides.

Using the information provided herein, such as specific Bordetella pathogenicity gene and polypeptide sequences, a polynucleotide of the invention encoding a Bordetella pathogenicity polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Bordetella pertussis* cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, typically a

library of clones of chromosomal DNA of *Bordetella pertussis* in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent hybridization conditions. By sequencing the individual clones thus identified by hybridization with sequencing primers designed from the original polypeptide or polynucleotide sequence it is then possible to extend the polynucleotide sequence in both directions to determine a full length gene sequence. Conveniently, such sequencing is performed, for example, using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Direct genomic DNA sequencing may also be performed to obtain a full length gene sequence.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than *Bordetella pertussis*, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions (for example, using a temperature in the range of 45 – 65°C and an SDS concentration from 0.1 – 1%) with a labeled or detectable probe consisting of or comprising a sequence defined in table 2 or 3 or a fragment thereof; and isolating a full-length gene and/or genomic clones containing said polynucleotide sequence.

The invention also provides a polynucleotide consisting of or comprising a polynucleotide sequence obtained by screening an appropriate library containing the complete gene for a polynucleotide sequence defined in tables 2 and 3 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence defined in table 2 or 3 or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers are described elsewhere herein.

The nucleotide sequence encoding Bordetella pathogenicity polypeptide encoded by the genes defined in tables 2 and 3 may be identical to the polypeptide encoding sequence contained in the genes defined in tables 2 or 3, or it may be a sequence, which as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide encoded by the genes defined in tables 2 and 3 respectively.

When the polynucleotides of the invention are used for the recombinant production of Bordetella pathogenicity polypeptide, the polynucleotide may include the coding sequence for the mature polypeptide or a fragment thereof, by itself; the coding sequence for the mature polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag, or is glutathione-s-transferase, or is MBP. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Nucleic acid comprising fragments of the sequences of the invention are also provided. These should comprise at least 10 consecutive nucleotides from the sequences (e.g. 12, 14, 15, 18, 20, 25, 30, 35, 40 or more depending on the particular sequence). Such fragments can preferably hybridise to the above-mentioned sequences under stringent conditions.

Further preferred embodiments are polynucleotides encoding Bordetella pathogenicity protein variants comprising the amino acid sequence of the Bordetella pathogenicity polypeptide encoded by the genes defined by tables 2 and 3 respectively in

which several, 10-25, 5-10, 1-5, 1-3, 1-2 or 1 amino acid residues are substituted, deleted or added, in any combination. Most preferred variant polynucleotides are those naturally occurring *Bordetella pertussis* sequences that encode allelic variants of the Bordetella pathogenicity proteins in *Bordetella* strains, preferably *B. pertussis*.

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 80%, and preferably at least 90%, and more preferably at least 95%, yet even more preferably 97-99% identity between the sequences.

Polynucleotides of the invention, which are identical or sufficiently identical to a nucleotide sequence of any gene defined in tables 2 and 3 or a fragment thereof, may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding Bordetella pathogenicity polypeptides respectively and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than *Bordetella pertussis*) that have a high sequence similarity to the Bordetella pathogenicity genes. Such hybridization techniques are known to those of skill in the art. Typically these nucleotide sequences are 80% identical, preferably 90% identical, more preferably 95% identical to that of the referent. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides. In one embodiment, to obtain a polynucleotide encoding Bordetella pathogenicity polypeptide, including homologs and orthologs from species other than *Bordetella pertussis*, comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having a nucleotide sequence contained in one of the gene sequences defined by tables 2 and 3, or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Thus in another aspect, Bordetella pathogenicity polynucleotides of the present invention further

include a nucleotide sequence comprising a nucleotide sequence that hybridize under stringent condition to a nucleotide sequence having a nucleotide sequence contained in one of the genes defined by table 2 and 3, or a fragment thereof. Also included with Bordetella pathogenicity polypeptides are polypeptides comprising amino acid sequences encoded by nucleotide sequences obtained by the above hybridization conditions. Such hybridization techniques are well known to those of skill in the art. Stringent hybridization conditions are as defined above or, alternatively, conditions under overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

A coding region of a Bordetella pathogenicity gene may be isolated by screening using a DNA sequence defined in table 2 or 3 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

There are several methods available and well known to those skilled in the art to obtain full-length DNAs, or extend short DNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman, *et al.*, *PNAS USA 85:* 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the "missing" 5' end of the DNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using "nested" primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5'

in the selected gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length DNA constructed either by joining the product directly to the existing DNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

The polynucleotides of the invention that are oligonucleotides derived from a sequence defined in table 2 or 3 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

Diagnostic Assays

This invention also relates to the use of Bordetella pathogenicity polypeptides, or Bordetella pathogenicity polynucleotides, for use as diagnostic reagents. Detection of Bordetella pathogenicity polypeptides will provide a diagnostic tool that can add to or define a diagnosis of *B. pertussis* disease, among others.

Materials for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy.

Thus in another aspect, the present invention relates to a diagonostic kit for a disease or suspectability to a disease, particularly *B. pertussis* disease, which comprises:
(a) a Bordetella pathogenicity polynucleotide, preferably the nucleotide sequence of one of the gene sequences defined by tables 2 and 3, or a fragment thereof ;
(b) a nucleotide sequence complementary to that of (a);

(c) a Bordetella pathogenicity polypeptide, preferably the polypeptide encoded by one of the gene sequences defined in tables 2 and 3, or a fragment thereof;

(d) an antibody to a Bordetella pathogenicity polypeptide, preferably to the polypeptide encoded by one of the gene sequences defined in tables 2 and 3; or

(e) a phage displaying an antibody to a Bordetella pathogenicity polypeptide, preferably to the polypeptide encoded by one of the gene sequences defined in tables 2 and 3.

It will be appreciated that in any such kit, (a), (b), (c), (d) or (e) may comprise a substantial component.

Polypeptides and polynucleotides for prognosis, diagnosis or other analysis may be obtained from a putatively infected and/or infected individual's bodily materials. Polynucleotides from any of these sources, particularly DNA or RNA, may be used directly for detection or may be amplified enzymatically by using PCR or any other amplification technique prior to analysis. RNA, particularly mRNA, cDNA and genomic DNA may also be used in the same ways. Using amplification, characterization of the species and strain of infectious or resident organism present in an individual, may be made by an analysis of the genotype of a selected polynucleotide of the organism. Deletions and insertions can be detected by a change in size of the amplified product in comparison to a genotype of a reference sequence selected from a related organism, preferably a different species of the same genus or a different strain of the same species. Point mutations can be identified by hybridizing amplified DNA to labeled Bordetella pathogenicity polynucleotide sequences. Perfectly or significantly matched sequences can be distinguished from imperfectly or more significantly mismatched duplexes by DNase or RNase digestion, for DNA or RNA respectively, or by detecting differences in melting temperatures or renaturation kinetics. Polynucleotide sequence differences may also be detected by alterations in the electrophoretic mobility of polynucleotide fragments in gels as compared to a reference sequence. This may be carried out with or without denaturing agents. Polynucleotide differences may also be detected by direct DNA or RNA sequencing. See, for example, Myers *et al.*, *Science, 230:* 1242 (1985). Sequence changes at specific locations also may be

revealed by nuclease protection assays, such as RNase, V1 and S1 protection assay or a chemical cleavage method. See, for example, Cotton *et al.*, *Proc. Natl. Acad. Sci., USA, 85:* 4397-4401 (1985).

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of a polynucleotide of the invention, which is associated with a disease or pathogenicity will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, a prognosis of a course of disease, a determination of a stage of disease, or a susceptibility to a disease, which results from under-expression, over-expression or altered expression of the polynucleotide. Organisms, particularly infectious organisms, carrying mutations in such polynucleotide may be detected at the polynucleotide level by a variety of techniques, such as those described elsewhere herein.

The invention further provides a process for diagnosing disease, preferably bacterial (particularly *Bordetella*) infections, more preferably infections caused by *Bordetella pertussis*, comprising determining from a sample derived from an individual, such as a bodily material, an increased level of expression of polynucleotide having a sequence defined in table 2 or 3. Increased or decreased expression of a polynucleotide can be measured using any on of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting, spectrometry and other hybridization methods.

Vectors, Host Cells, Expression Systems

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

Recombinant polypeptides of the present invention may be prepared by processes well known in those skilled in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems that comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems, and to the production of polypeptides of the invention by recombinant techniques.

For recombinant production of the polypeptides of the invention, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis, *et al.*, *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook, *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as cells of streptococci, staphylococci, enterococci, *E. coli*, streptomyces, cyanobacteria, *Bacillus subtilis*, *Moraxella catarrhalis*, *Haemophilus influenzae* and *Neisseria meningitidis*; fungal cells, such as cells of a yeast, *Kluveromyces*, *Saccharomyces*, a basidiomycete, *Candida albicans* and *Aspergillus*; insect cells such as cells of *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293, CV-1 and Bowes melanoma cells; and plant cells, such as cells of a gymnosperm or angiosperm.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal-, episomal- and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia

viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, picornaviruses, retroviruses, and alphaviruses and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

In recombinant expression systems in eukaryotes, for secretion of a translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, ion metal affinity chromatography (IMAC) is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and or purification.

The expression system may also be a recombinant live microorganism, such as a virus or bacterium. The gene of interest can be inserted into the genome of a live recombinant virus or bacterium. Inoculation and *in vivo* infection with this live vector will lead to *in vivo* expression of the antigen and induction of immune responses. Viruses and bacteria used for this purpose are for instance: poxviruses (e.g; vaccinia,

fowlpox, canarypox), alphaviruses (Sindbis virus, Semliki Forest Virus, Venezuelian Equine Encephalitis Virus), adenoviruses, adeno-associated virus, picornaviruses (poliovirus, rhinovirus), herpesviruses (varicella zoster virus, etc), Listeria, Salmonella , Shigella, Neisseria, BCG. These viruses and bacteria can be virulent, or attenuated in various ways in order to obtain live vaccines. Such live vaccines also form part of the invention.

Antibodies

According to a further aspect, the invention provides antibodies which bind specifically to the polypeptides of the invention. These may be polyclonal or monoclonal and may be produced by any suitable means well known to a skilled person in the art.

Typically, a mouse or rat is immunised with a protein (preferably adjuvanted with Freund's complete adjuvant) and injected (doses of 50-200 µg/injection is typically sufficient). Polyclonal antibodies can be isolated by bleeding the animal to extract serum. Alternatively, monoclonal antibodies can be generated by removing the spleen (or large lymph nodes) and dissociating it into single cells (Kohler and Milstein, (1975) Nature, 256:495-497). These are then induced to fuse with myeloma cells to form hybridoma, and are cultured in a selective medium (eg hypoxanthine, aminopterin, thymidine merium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected monoclonal-secreting hybridomas are then cultured either in vitro (eg in tissue culture bottles or hollow fiber reactors), or in vivo (as Ascites in mice).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides or polynucleotides of this invention. Also, transgenic mice, or other organisms or animals, such as other mammals, may be used to express humanized antibodies immunospecific to the polypeptides or polynucleotides of the invention.

Alternatively, phage display technology may be utilized to select antibody genes with binding activities towards a polypeptide of the invention either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-Bordetella pathogenicity polypeptide or from naive libraries (McCafferty, *et al.*, (1990), Nature 348, 552-554; Marks, *et al.*, (1992) *Biotechnology 10*, 779-783). The affinity of these antibodies can also be improved by, for example, chain shuffling (Clackson *et al.*, (1991) *Nature 352:* 628).

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides or polynucleotides of the invention to purify the polypeptides or polynucleotides by, for example, affinity chromatography.

Antibodies against a Bordetella pathogenicity polypeptide or polynucleotide may be employed to treat infections, particularly bacterial infections.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants form a particular aspect of this invention.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized," where the complimentarity determining region or regions of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones *et al.* (1986), *Nature* 321, 522-525 or Tempest *et al.*, (1991) *Biotechnology* 9, 266-273.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with Bordetella pathogenicity polypeptide or epitope-bearing fragments, analogs, outer-

membrane vesicles or cells (attenuated or otherwise) adequate to produce antibody and/or T cell immune response to protect said animal from Bordetella (particularly *B. pertussis*) disease, among others. Such agents may be used alone, or conjugated to another molecule which improves its immunological potency. In particular the invention relates to the use of Bordetella pathogenicity polypeptides encoded by the genes defined in table 3 – the effector proteins. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering Bordetella pathogenicity polypeptide via a vector directing expression of Bordetella pathogenicity polynucleotide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

A further aspect of the invention relates to an immunological composition or vaccine formulation which, when introduced into a mammalian host, induces an immunological response in that mammal to a Bordetella pathogenicity polypeptide (particularly one encoded by a gene defined in table 3) wherein the composition comprises a Bordetella pathogenicity gene, or Bordetella pathogenicity polypeptide or epitope-bearing fragments, analogs, outer-membrane vesicles or cells (attenuated or otherwise). The vaccine formulation may further comprise a suitable carrier. The Bordetella pathogenicity polypeptide vaccine composition is preferably administered orally or parenterally (including subcutaneous, intramuscular, intravenous, intradermal etc. injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems

known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

The vaccine formulations of the invention may also comprise other *Bordetella* antigens known to be suitable vaccinal agents, for instance: pertussis toxoid, pertactin, agglutinogins 1 and 2, FHA (filamentous haemagglutinin), and adenylate cyclase / haemolysin (AC/HLY), or immunogenic fragments thereof (Locht *et al.*, NAR (1986) 14:3251-3261; Relman *et al.*, PNAS USA (1989) 86:2637-2641; Roberts *et al.*, Mol. Microbiol. (1991) 5:1393-1404; Mooi *et al.*, Microb. Pathog. (1992) 12:127-135; Hewlett and Gordon, In *Pathogenesis and Immunity in Pertussis* (1988), New York, Wiley & Sons, pp. 193-209.

Yet another aspect of the invention relates to an immunological/vaccine formulation which comprises the polynucleotide of the invention. Such techniques are known in the art, see for example Wolff *et al.*, *Science*, (1990) *247*: 1465-8.

Vaccine compositions can comprise polypeptides, antibodies, or polynucleotides of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition (in this case *Bordetella*, particularly *B. pertussis*, disease), or to exhibit a detectable therepeutic or preventative effect. The effect can be detected by, for example, antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention.

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples illustrate, but do not limit the invention.

Example 1: A type III secretion system is present in a pathogenicity island in *Bordetella pertussis*.

The presence of a *lcrD* homologous gene in the *Bordetella pertussis* genome was investigated by polymerase chain reaction (PCR). The primers used (oligos 95080 and 95081 shown in Table 1) were degenerate oligonucleotides corresponding to highly conserved regions of the amino acids sequences of the LcrD/FlbF family of proteins. These primers were also designed to favour the amplification of virulence genes instead of their paralogue *flhA* or *flbF* flagellar genes, present in flagellated bacterial strains. The presence of the 3' triplet CAT in oligonucleotide 95081 is a determinant – indeed when multiple sequence analysis is done using known homologous sequences (database searching was done with either the FASTA and TFASTA programs of the GCG9 package, or with BLASTN, BLASTP and BLASTX programs, and alignments were carried out with the PILEUP program from the GCG9 package) it could be seen that the CAT triplet codes for a methionine which is exclusively present in virulence sequences while absent in the flagellar ones.

When analysed on agarose gel, the PCR product appeared as a heterogeneous mix of fragments, one of which was presenting the expected size (around 150 bp). A second round of amplification using the approximately 150 bp DNA as template yielded a single amplicon which was cloned in pCRII (obtained from Invitrogen) for further characterisation. It appeared as a 152 bp fragment whose nucleotide sequence (Fig. 1), although similar to all *lcrD/flbF* homologous genes, shares a higher level of identity with the virulence (*lcrD*-like) genes.

Table 1.

oligonucleotides	sequence[1]	features	*lcrD* corresponding codons[2]
95080	GSH ATG CCW GGH AAR CAR ATG	direct, degenerate	150 to 156
95081	GC RTC DCC YTT DAC RAA YTT CAT	complement, degenerate	193 to 200
95363	CC ATC GAC GCG GAC TTG CGC G	direct, non-degenerate	157 to 164
95364	CGC GCC GTC CAT GGC GCC ATA	complement, non-degenerate	186 to 192
96110	C CGA CGC CGA CGC CGT ACG GTC	direct, non-degenerate	172 to 179

[1] The letter code for nucleotide ambiguity proposed by IUB (Nomenclature Committee, 1985, Eur. J. Biochem., **150**: 1-5) was used.

[2] The DNA sequence of the *lcrD* gene from *Yersinia enterocolitica* used for this work was published by Plano *et al.* (1991).

To ensure that the cloned fragment was actually a *B. pertussis* sequence PCR was performed under stringent conditions with serial 10-fold dilutions of DNA from *B. pertussis*. The optimisation of stringent PCR conditions require a perfect match between template and primers. It was likely, however, that due to the degeneration of the original primers, the 152 bp sequence initially obtained had, at its boundaries, a few base pair differences with the actual *B. pertussis lcrD*-like (hereafter called *bcrD*) sequence. A nested PCR approach using internal primers (oligos 95363 and 95364 Table 1) was therefore preferred, as primers known to be the correct *B. pertussis* sequence are used. A dose-response-relationship was observed between the 10-fold dilutions of *B. pertussis* template DNA and the product of the nested PCR, suggesting that the 152 bp amplicon actually originates from the *Bordetella* genome.

Comparison of the 152 bp sequence with *lcrD/flbF* genes allowed us to define a specific DNA stretch (oligo 96110 in Table 1) which was used as a probe for screening a

genomic library of *B. pertussis* constructed in the plasmid vector pBR327 (Delisse-Gathoye *et al*, 1990, *Infect-Immun.* **58**: 2895-905). Several positive clones were isolated and restriction analysis of their resident plasmids showed that they harboured overlapping inserts. The entire nucleotide sequence of one insert was determined, revealing a large open reading frame (ORF). This 2100 bp ORF encoded a 75 kDa polypeptide which is 59 % and 47 % identical to the yersinial proteins LcrD and FlhA respectively. Multiple amino acids comparisons of all known members of the LcrD/FlbF family of proteins, including the *B. pertussis* BcrD deduced amino acid sequence, showed that this sequence clearly ranked within the virulence associated determinants (Fig. 2). These data strongly suggest that *B. pertussis* possesses a type III export system, involved in the secretion of virulence effectors.

The *B. pertussis lcrD*-like nucleotide sequence (*bcrD*) has been submitted to EMBL and assigned the accession number Y13383.

This general technique has been useful for determining the presence/absence of a type III secretion system in other bacterial strains. The human pathogens *Borrelia burgdorferi* and *Helicobacter pylori* were intensively screened for such a system using this technique. No evidence for a type III secretion system could be found. The subsequent publication of the genome sequences of these microorganisms has confirmed the absence of similar systems in these species. In contrast, the method allowed the amplification of a DNA fragment from the phytopathogen *Pseudomonas corrugata*, which clearly ranks among the virulence sequences. This technique could be applied to any Gram negative pathogen of medical or agronomic importance such as *Neisseria* spp, *Moraxella catharalis*, *Vibrio cholerae*, any Enterobacteriaceae, *Pseudomonas* spp, *Haemophilus influenzae*, *Brucella* spp, *Francisella tularensis*, *Pasteurella* spp, *Legionella pneumophila*. Even in strains that have been fully sequenced, this technique can be used as a simple method for checking alternate types or strains of the same species. For instance, some types of pathogenic *Escherichia coli* harbour a type III secretion system whereas others do not.

Example 2: Analysis of the *B. pertussis bcrD* flanking sequences to characterise the pathogenicity island and virulence-related proteins encoded therein

The tendency for systematic clustering of type III encoding genes inside pathogenicity islands prompted the analysis of *B. pertussis bcrD* flanking sequences. The whole region containing the pathogenicity island was sequenced by chromosome walking taking care to pay attention to the fact that each Pathogenicity island region must be represented in at least two independent clones, to avoid possible artefacts due to chimeric DNA inserts. This revealed clustered ORFs that could be classed in 3 categories: class I type ORFs (table 2); class II type ORFs (table 3) – the effector proteins which have the best vaccinal and diagnostic properties; & insertion sequences, and ORFs homologous to house keeping genes of other species (table 4). Although there is no general rule for defining the boundaries of a Pathogenicity island, they can be demarcated with a direct or inverse repeat at one or other boundary, however the absolute demarcation of the boundaries can only really be done by the detection of house keeping genes at the extremes of the sequence. In the present case, an insertion sequence (IS in Fig. 3) was present at the 5' end of the island (separating the virulence ORFs from the house keeping genes), but absent at the 3' end. In addition, the presence of house keeping genes (*greA* and ICFG-like) surrounding a locus which, according to sequence data, encompasses numerous virulence sequences is a good indication of the boundaries of the island. The complete gene organisation of the pathogenicity island is schematically represented in Figure 3. The precise definition of the PAI boundaries requires further experimental data, such as the characterisation of the corresponding chromosomal region of a *Bordetella* strain which is devoid of a type III secretion system.

Table 2

names	Coding sequence from/to (with reference to Fig. 5)	Coding DNA strand	SEQ ID NO:	Homologous genes (from *Yersinia*, unless otherwise specified)
Class I genes, i.e. genes coding for determinants involved in the secretory apparatus and their regulation				
bcrD	8656/10755	complement	1	*LcrD*
bcrH	14097/14582	direct	3	*lcrH* (= *sycD*)
bscC	26955/28757	direct	5	*YscC*
bscD	7379/8659	complement	7	*YscD*
bscE	7039/7338	complement	9	None
bscF	6783/7049	complement	11	*YscF*
bscI	17892/18218	direct	13	*YscI*
bscJ	18215/19039	direct	15	*YscJ*
bscK	19032/19694	direct	17	None
bscL	19664/20302	direct	19	*YscL*
bscN	20307/21641	direct	21	*YscN*
bscO	21641/22150	direct	23	*YscO*
bscP	22147/22695	direct	25	None
bscQ	22692/23771	direct	27	*YscQ*
bscR	23768/24439	direct	29	*YscR*
bscS	24445/24711	direct	31	*YscS*
bscT	24723/25523	direct	33	*YscT*
bscU	25520/26569	direct	35	*YscU*
bscV	26566/26964	direct	37	None
brpL	28778/29380	complement	39	*hrpL* (*Pseudomonas syringae*)

Table 3

Names	Coding sequence from/to (with reference to Fig. 5)	Coding DNA strand	SEQ ID NO:	Homologous genes (from *Yersinia*, unless otherwise specified)
Class II ORFs which putatively code for effector proteins				
bopN	11906/13003	complement	41	*YopN* (= *lcrE*)
orf1	6160/6747	direct	43	None
orf2	10752/11120	complement	45	None
orf3	11117/11527	complement	47	None
orf4	11532/11909	complement	49	None
orf5	13002/13784	direct	51	None
orf6	13806/14081	direct	53	None
orf7	14630/15571	direct	55	None
orf8	15601/16803	direct	57	None
orf9	16827/17288	direct	59	*BcrH*
orf10	17293/17814	direct	61	*pcr4* (*Pseudomonas aeruginosa*)
orf11	29412/29591	complement	63	None
orf12	29555/30529	complement	65	None
orf13	30631/31776	direct	67	None
orf14	31773/33005	complement	69	None
orf15	32370/33014	direct	71	None

Table 4

No name specified	Coding sequence from/to (with reference to Fig. 5)	Coding DNA strand	SEQ ID NO:	Homolgous sequences
Insertion Sequences and house keeping genes				
	711/2024	direct	73	uracil permease genes of numerous bacteria
	2055/3590	complement	75	Chemoreceptor genes of numerous bacteria
	4220/4696	direct	77	*greA* (*Escherichia coli*)
	4998/5948	complement	79	transposase genes of numerous bacteria
	33002/34852	complement	81	ICFG gene (*Synechocystis* sp)

Next to the *bcrD* gene, there is an open reading frame (ORF) whose deduced amino acid sequence shares significant similarities with the YscU protein of *Yersinia* spp (39% identity and 51% similarity) and other known YscU homologs (Fig. 4). YscU, like LcrD, is a component of the *Yersinia* type III secretion machinery involved in the virulence mechanisms of the bacteria. *B. pertussis* therefore possesses a classical type III secretion system which is most probably involved in pathogenicity. This latter point can be investigated through phentoypic analyses of mutants (see below).

The total length of the Pai is approximately 30 to 40 kb. The DNA sequence of the whole region is presented in Figure 5, and is referred to in tables 2, 3, and 4. Restriction analysis on pulsed-field gel electrophoresis allowed the type III locus to be mapped at coordinate position 1,590 kb on the Tohama I strain chromosome.

No homologies could be found between the *B. pertussis* Class II Pai DNA sequences and the sequences reported in the GenEMBL databases (except for those stated in table 3). The expressed products of these unknown genes within the Pai

responsible for virulence, will be useful in the development of a vaccine formulation against pathogenic *Bordetella pertussis*.

To address the precise function of the Pai, a *bcrD* mutant was engineered by allelic exchange. In the resulting mutant, the *bcrD* gene was disrupted by an *aphA*-3 cassette conferring kanamycin resistance. This cassette was inserted in such a way that translation was not interrupted, avoiding any polar effect on expression of putative downstream cistrons. A mutant has been isolated and its associated phenotype is being currently analysed.

Example 3: Analysis of the *in situ* expression of the genes of the pathogenicity island

Genetic constructions

To produce a mutant defective in type III secretion, a 255-bp fragment (codons 363 to 445) was deleted from the *bcrD* coding sequence and replaced by a cassette containing the *aphA-3* gene which confers kanamycin resistance (Menard *et al.*, J. Bacteriol. (1993) 175:5899-5906). The *aphA-3* cassette was excised from pUC18K by *Eco*RI-*Pst*I digestion and introduced in the *bcrD Eco*RI-*Sse*8387I sites. This construct generated an early stop in *bcrD* translation and allowed in-frame translation of the remaining 3' end of the mutated gene, avoiding possible polar effects on expression of downstream cistrons. The mutated *bcrD* gene with its flanking sequences, was excised by *Bgl*II-*Not*I cutting and subsequently inserted into the *Xba*I-*Eco*RI sites of the suicide plasmid pSS1129 (Stibitz, Methods Enzymol. (1994) 235:458-465), thanks to DNA adaptators. The resulting construction was named pAF214. pAF248 is a derivative of pAF214 that contained two additional unique *Spe*I and *Pac*I sites. These sites, included in a pair of complementary oligonucleotides, were introduced into the *Bam*HI site of pAF214. Other constructs included pAF245 and pAF246. PCR amplification of a 831 bp fragment covering the 5' region and the 4 first codons of *bcrD* was generated. This amplicon was further introduced into *Bam*HI-*Hin*DIII linearized pNM480 (Minton, Gene (1984) 31:269-273), in such a way that the *bcrD* initiation codon was placed in frame

with *lacZ*, used as a reporter gene. The resulting construct was named pAF245. Similarly, primers were designed for placing *lacZ* downstream of a 849 bp fragment that encompassed upstream *bscN* sequences including its 3 first codons. pAF246 was obtained by cloning this fragment in pNM480.

Transformations and allelic exchanges

B. pertussis cells, from a freshly saturated culture in 10 ml of SS medium, were washed and resuspended in 100µl of a cold 10% (v/v) glycerol solution. Up to 10 µg of supercoiled purified DNA in a maximum of 20 µl of water were added to 100 µl of the bacterial suspension. Cells and DNA were transferred to a prechilled 0.2 cm electroporation cuvette (Bio-Rad) and placed in a Gene Pulser apparatus (Bio-Rad). Pulses were achieved with settings of 25 µF, 2.5 kV, and 600 Ω, giving a time constant ranging from 11 to 14 ms.

After their initial isolation on BG plus gentamycin, pAF214 and pAF248 transformants that undergone a second recombination step were selected on streptomycin as described (Stibitz, *supra*). The null *bcrD* mutants were finally distinguished from revertants by their acquired resistance to kanamycin. The proper integration of the *aphA-3* was assessed by southern blot analysis. In contrast, introduction of pAF245 and pAF246 only required a single crossover selected on BG plus ampicillin. This recombination step led to the placement of the *lacZ* coding sequence under the control of the signals governing the transcription of *bcrD* and *bscN* respectively.

Mice model

After a two days growing on BG agar plates, wild type and mutant bacteria were recovered and resuspended in PBS at a concentration of 10^8 PFU ml^{-1}. 25 µl of the suspension were injected in each nostril of pentobarbital anaesthetized mice. Lungs colonization was assayed after 4 h, 3, 7, 14, 26, 39 and 45 days by treating both lungs of each mouse in an Ultraturax grinder and titrating the resuspended bacteria on BG agar plates.

β-galactosidase assay

0.5 ml of bacterial suspensions coming from liquid cultures grown to log phase (OD = 0.2), were assayed as described previously (Miller, (1972) "Experiments in molecular genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.). We used the chromogenic substrate *o*-nitrophenyl-*β*-D-galactoside (ONPG) of Sigma.

Transcription of both bcrD *and* bscN *transcripts appear controlled by the* bvg *locus*

Most of the *Bordetella* virulence functions are controlled by the *bvg* locus. The Bvg^+ phase is characterized by the expression of virulence factors and is necessary for colonization of animal models. In contrast, the bacteria are avirulent in Bvg^- phase which can be induced by nicotinic acid or $MgSO_4$. We investigated the level of expression of two genes that belonged to distinct unit of transcription, i.e. *bcrD* and *bscN*, by using transcriptional fusions of *lacZ* into these genes. To this end, we isolated the mutants NIVh86 and NIVh87, which integrated pAF245 and pAF246 respectively. In the former mutant, a single recombination step led to the setting *lacZ* in place of the *bcrD* coding sequence, whereas in the latter, *lacZ* replaced *bscN*. The level of expression of both *bcrD* and *bscN* transcripts was assessed either in Bvg^+ or in Bvg^- phases. Both *B. pertussis* genes were weakly expressed *in vitro*. Additionally, however, these levels of expression appeared to be clearly modulated by the Bvg system. Indeed, whereas β-galactosidase could be assayed in Bvg^- conditions, no enzyme activity was detected in Bvg^- phase (table 5).

Table 5. β-galactosidase activity, in Miller units (Miller, *supra*), when *lacZ* is placed under the control of that direct the exrpession of *bcrD* or *bscN*.

phase transcript	Bvg^+	Bvg^-
bcrD	3.54	0.02
bscN	1.65	0.04

Example 4: Recombinant expression of effector protein vaccine candidates

In the discovered sequence, seven ORFs (*orf2* to *-8*) particularly fulfil certain criteria that make them good candidates as effector proteins and vaccine candidates. First, they appear surrounded by typical type III secretion (class I) genes, and therefore incontestably belong to the type III secretion locus. Furthermore, they don't display significant similarities with genes present in related type III systems from other organisms, and are therefore likely to be effector proteins specific for *Bordetella*. In addition to these ORFs *bopN*, *orf9* and *orf10* are also of particular interest as vaccine candidates. Despite the fact that these sequences do not fulfil the second criterium above (they have some similarity to *popN*, *pcrH* and *pcr4* of *Pseudomonas aeruginosa*), these products may also be exported by the specialized translocon. For these reasons, ten ORFs, i.e. *orf2* to *-10* and *bopN*, were selected for further analysis. To this end, ten pairs of primers (table 6) were designed for amplifying their corresponding ORF. The amplified ORFs were then cloned in the pCR-TOPO® T/A cloning system (Invitrogen) and their sequences were checked for errors putatively induced by the Taq DNA polymerase. Correct inserts were retrieved by *Eco*RI and *Bam*HI (or *Bgl*II - see table 6) cutting and transferred into the pMAL® vectors (New England Biolabs; Maina *et al.*, Gene (1988) 74:365-373), opened by *Eco*RI and *Bam*HI restriction. In these vectors, expression of the cloned inserts yields recombinant proteins fused to the maltose binding

protein (MBP) of *E. coli*. The MBP domain of the fusion protein provides a means for both detecting the expressed product and purifying it by affinity chromatography.

Four ORFs, namely *orf2*, *-4* and *-10* on the one hand, and *orf6* on the other, have been cloned into pMAL-c2E ® and pMAL-p2E® respectively. Transformed bacteria, grown in 300 ml of culture medium, were induced with IPTG (300 µM) and lysed in a French pressure cell. Insoluble material was pelleted by ultracentrifugation and discarded whereas the resulting supernatant was applied to an amylose resin. Fusion proteins that specifically bind to the amylose through their MBP domain, were further eluted by application of maltose 10 mM. This method allowed us to recover from 10 to 50 mg of each fusion protein (Fig. 6). The expressed *Bordetella* products may be separated from the MBP by utilising the enterokinase cleavage site between the *Bordetella* polypeptide and the MBP. The other ORFs should be expressable using a similar approach.

The secreted proteins will be analysed using standard techniques to confirm their functional and immunological properties. First, the immunogenicity of the secreted proteins will be assessed by investigating the presence of antibodies directed against these proteins in the serum of infected patients. In addition, their putative recognition as protective antigens will be based on challenge experiments, realized in a mouse model. Second, the biological properties of the effector proteins will be assessed by analysing their catalytic activities. For instance, it is expected that one of the secreted proteins would display a tyrosine phosphatase activity. Finally, the function of the effector proteins will be investigated by microinjecting the proteins into the cytoplasm of eukaryotic cells. This will allow us to display putative activities of inhibition of actin polymerisation, cytotoxicity or induction of apoptosis, i.e. those types of activities that have been assigned to effector proteins secreted by type III secretion systems discovered in other species.

Table 6. PCR primers used for amplifying the ORFs encoding vaccine candidates.

orf2	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG ATG CCG CAT ACC CTA CCC TCG
	complement	5'-TCT AGA GGA TCC GGC GAA TGG ATT TCT TGC TCG TCA
orf3	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG TCC AGC GCC GTA CCC GGC
	complement	5'-TCT AGA GGA TCC AGG GTA GGG TAT GCG GCA TCA TCC
orf4	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG AAT ACT GCC GAT AGG GCG CTG
	complement	5'-TCT AGA GGA TCC GGT ACG GCG CTG GAC ATG GCG TC
bopN	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG ACT CGT ATC GAT GCC GCC
	complement	5'-TCT AGA GGA TCC GCG CCC TAT CGG CAG TAT TCA TGC
orf5	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG GGG AGT CCT CGG AGA AGG AA
	complement	5'-TCT AGA GGA TCC ATA CTC CTT GTG CAG CGC TTA GCG
orf6	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG CAG GAG CAA GGC ATC CAA TC
	complement	5'-TCT AGA GGA TCC CAT GGA AGG CCT CCG CGC TCA GAC
orf7	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG TCT GTT TCT CCG ACT TCG CCC
	complement	5'-TCT AGA GGA TCC TGA AGG TTG GAG CCG GAC ACT CAG
orf8	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG ACC GTC ATG AGT ACG ACC ATA
	complement	5'-TCT AGA TCT TTC CTT GAG CGC CCG GCG CTA CA
orf9	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG ACT GTT CAT GAC GAC GCG
	complement	5'-TCT AGA GGA TCC GAG TCT GAG TGC ATG GAG TTA CTC C
orf10	direct	5'-GAG GAA TTC CAT ATG CCC ACC ATG CAC TCA GAC TCA GGT TCA GAT
	complement	5'-TCT AGA GGA TCC TCG CCG TCA GAT CCA AAT TCA TCC AG

Initiation and STOP codons of the corresponding ORF are written in bold. The cloning sites *Eco*RI, *Bam*HI or *Bgl*II are underlined. All but one of the complementary primers contain a *Bam*HI site. In the case of *orf8*, as it presents an internal *Bam*HI recognition sequence, a *Bgl*II site was preferred.

Bibliography

Allaoui, A., Woestyn, S., Sluiters, C., and Cornelis, G. R. (1994) YscU, A Yersinia enterocolitica inner membrane protein involved in Yop secretion. *J Bacteriol* **176**:4534-4542.

Bergman, T., Erickson, K., Galyov, E., Persson, C., and Wolf-Watz, H. (1994) The lcrB (yscN/U) gene cluster of Yersinia pseudotuberculosis is involved in Yop secretion and shows high homology to the spa gene clusters of Shigella flexneri and Salmonella typhimurium. *J Bacteriol* **176**:2619-2626.

Bogdanove, A. J., Wei, Z.-M., Zhao, L., and Beer, S. V. *Erwinia amylovora* secretes a Harpin via a type III pathway and contains a homolog of *yopN* of *Yersinia* spp. *J Bacteriol* **178**:1720-1730.

Fenselau, S., Balbo, I., and Bonas, U. (1992) Determinants of pathogenicity in Xanthomonas campestris pv. vesicatoria are related to proteins involved in secretion in bacterial pathogens of animals. *Mol Plant Interact* **5**:390-396.

Gyri (1995). *Mol. Microbiol.* **15**: 761-769.

Hueck, C. J. (1998) Type III protein secretion systems in Bacterial Pathogens of Animals and Plants. *Microb. Mol. Biol. Rev.* **62**: 379-433

Plano (1991). *J. Bact.* **173**: 7293-7303.

Ramakrishnan, G., Zhao, J. L., and Newton, A. (1991) The cell cycle-regulated flagellar gene flbF of Caulobacter crescentus is homologous to a virulence locus (lcrD) of Yersinia pestis. *J Bacteriol* **173**:7283-7292.

Yuk, M. H., Harvill, E. T., and Miller, J. F. (1998) The BvgAS virulence control system regulates type III secretion in *Bordetella bronchiseptica*. *Mol. Microbiol.* **28**:945-959.